PCT/AU00/01038

REC'D 26 SEP 2000	
WIPO	PCT

4

Patent Office
Canberra

AU00/01038

I, KAY WARD, ACTING MANAGER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 2612 for a patent by THE BIONIC EAR INSTITUTE filed on 02 September 1999.

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

WITNESS my hand this
Twenty-first day of September 2000

K Ward

KAY WARD
ACTING MANAGER EXAMINATION
SUPPORT AND SALES

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

P/00/009
Regulation 3.2

AUSTRALIA
Patents Act 1990

PROVISIONAL SPECIFICATION

FOR THE INVENTION ENTITLED:

"IMPROVED SOUND PROCESSOR FOR COCHLEAR IMPLANTS"

Applicant:

THE BIONIC EAR INSTITUTE

The invention is described in the following statement:

IMPROVED SOUND PROCESSOR FOR COCHLEAR IMPLANTS

Field of the Invention

This invention relates to improvements in sound processors for cochlear implants.

Background of the Invention

The multi-channel cochlear implant was first implanted in 1978. Early signal processing designs extracted the second formant (F2) and pitch (F0) to control electrode stimulation. The frequency of F2 controlled the location of electrode stimulation, and F0 controlled the rate of stimulation. This was later improved by also extracting the first formant (F1) and adding a second stimulated electrode for each pitch period. The MULTI-PEAK (MPEAK) stimulation strategy added stimulation of a number of fixed electrodes to better represent high-frequency information. The next stages of development were the SMSP and SPEAK strategies. These were a departure from the others at they used a fixed stimulation rate and stimulated electrodes that corresponded to maxima in the sound spectra. Another fixed-rate strategy, CIS, was developed overseas. This strategy stimulated all of a small number of electrodes to represent the sound spectra. All of the above processing strategies involve fixed-rate sound processing.

It has been determined that some speech features are better perceived using low-rates of simulation, while some are better perceived using high-rates of stimulation. The fixed rate of stimulation used in the past is a trade-off between the transmission of these features. While higher rates of stimulation present more information about phonetic manner of articulation, the refractory properties of the auditory nerve cause spectral information to be smeared at such higher rates.

Summary of the Invention and Object

It is an object of the present invention to provide an improved sound processor for use with cochlear implants in which the problems associated with fixed rate stimulation are ameliorated.

The invention provides an improved sound processor for a cochlear implant having electrodes for stimulating the auditory nerve, including means for receiving

sounds, means for processing the sounds and converting them to electrical stimulation signals for application to the electrodes of the cochlear implant for stimulation of the auditory nerve, said sound processing means including means for generating electrical signals to be applied to the basal electrodes having different predetermined rates of stimulation.

In one form of the invention, the cochlear implant has basal electrodes and apical electrodes and the means for generating electrical signals to be applied to the apical electrodes have a different rate of stimulation, the electrical signals to be applied to the basal electrodes having a higher rate of stimulation than the electrical signals to be applied to the apical electrodes.

By causing stimulation of the basal electrodes at a higher rate of stimulation than the apical electrodes, the features of speech in the sounds being processed will be more optimally presented to the cochlear implant user, leading to improved speech understanding performance. The low rates of stimulation of the apical electrodes will present good spectral information in this region, where it is most important. High rates of stimulation at the basal electrodes will present good information about temporal events and frication.

In a preferred embodiment, the more apical electrodes will be chosen as those that contain the voice bar and lower formants of speech. In this frequency region, spectral detail is important and the apical electrodes will be stimulating using a stimulation rate of between about 250 cycles per second and about 800 cycles per second, depending on the user. By adopting stimulation rates falling within the above range, better information about place of articulation of speech, which is largely represented by the formant structure, is obtained by the user.

The more basal electrodes represent higher frequency components of the incoming sound, and higher rates of stimulation of these electrodes will be used to better represent noise and more precisely present information about temporal events such as rapid changes in amplitude. The latter is important for perception of manner of articulation and voicing. These electrodes will be stimulated at a higher rate than the apical electrodes, with stimulation rates between about 800 cycles per second and about 1600 cycles per second being selected depending on the user.

In the case of an implant having 20 electrodes available for stimulation, the apical electrodes are electrodes 0 to 12, and the basal electrodes are electrodes 13 to 19. The apical electrodes represent sound frequencies from 0 to about 2700Hz, while the basal electrodes represent frequencies from about 2700Hz to about 7900Hz. The stated apical electrode frequencies are sufficient to contain the first three formants of most speakers speech.

In a particularly preferred form of the invention, the apical electrodes are stimulated at about 250 cycles per second while the basal electrodes are stimulated at about 1600 cycles per second. To ensure that stimulation levels are suitable for these different rates, the threshold (T) levels and comfort (C) levels of the patient are carefully set. The electrodes to be stimulated are chosen by selecting the eight largest spectral energies within filterbanks derived from the Fast Fourier Transform (FFT) or the Discrete Wavelet Transform (DWT) which is performed by the processor.

In another form the invention provides an improved sound processor for a cochlear implant having electrodes for stimulating the auditory nerve, including means for receiving sounds, means for processing the sounds and converting to electrical stimulation signals for application to the electrodes of the cochlear implant whereby the auditory nerve is electrically stimulated, said sound processing means having means for varying the rate of stimulation of the electrical stimulation signals depending on the parameters of the sound received by the sound receiving means.

By varying the rate of stimulation of the cochlear implant electrodes depending on the incoming speech signal, key speech features will be more optimally presented to the cochlear implant user thereby leading to improve speech understanding performance.

In a preferred form of this aspect the invention, the sound processing means will be programmed to continually adjust the rate of stimulation of the electrical stimulation signals depending on the parameters of the incoming speech signal. To this end, the incoming speech signal will be processed to detect events that are better represented using a higher rate of stimulation. Such events include plosive

onset bursts, frication and rapid spectral changes. The rate of stimulation across all electrodes will be increased for the average duration of these events. The standard rate will be between 250 cycles/s and 800 cycles/s depending on the user. The higher rate will be between 800 cycles/s and 1600 cycles/s, also depending on the user.

In order that the invention may be more readily understood, one presently preferred embodiment of the invention will now be described.

The invention is preferably designed for use with the CI 24M Cochlear Implant as manufactured by Cochlear Ltd, and as described in US Patent No. 4532930, the contents of which are incorporated herein by cross-reference, and later patents by Cochlear Ltd to be found in the patent literature.

Although the CI 24M Implant will be used in most cases, the invention could be applied to any implant that uses pulsatile stimulation. The stimulation strategy is based on the Spectral Maxima Sound Processor (SMSP), which is described in United States Patent 5597380 and Australian Patent 657959, although other strategies may be used with similar results. The electrode selection strategy from the SMSP is varied to ensure that electrodes are stimulated at the desired predetermined frequencies for each cycle of stimulation. The preferred signal processing device will be the SPEAR processor, which is currently under development at The Bionic Ear Institute, and which is described in the following paper :

Zakis, J.A. and McDermott, H.J. (1999). "A new digital sound processor for hearing research," Proceedings of the Inaugural Conference of the Victorian Conference of the Victorian Chapter of the IEEE Engineering in Medicine and Biology Society, February 22-23, pp. 54-57.

The processor is a generic processor based on the Motorola DSP56302, although any digital signal processor, including those produced by Cochlear Ltd and their competitors, could be used to run the differential rate sound processor program, provided they have adequate processing speed.

In the implementation of the first form of the invention, the differential rate stimulation processor software embodying the invention is downloaded to the

SPEAR processor and stored on an EPROM. Patient map details, including frequency bands, threshold (T) levels and comfort (C) levels, are also stored on the device. Monopolar stimulation mode is used to reduce current levels and for longer battery life.

For the case where 20 electrodes are available for stimulation, the apical electrodes are electrodes 0 to 12, and the basal electrodes are electrodes 13 to 19. The apical electrodes then represent frequencies from 0 to 2700Hz; the basal electrodes represent frequencies from 2700Hz to 7900Hz. The stated apical electrode frequencies are sufficient to contain the first three formants of most speakers' speech.

The apical electrodes are stimulated at about 250 cycles/s and the basal electrodes at about 1500 cycles/s. The patient's T and C levels are carefully set to ensure that stimulation levels are suitable for the two different rates and adjustments made if necessary. The electrodes to be stimulated are chosen by selecting the eight largest spectral energies within filterbanks derived from the Fast Fourier Transform (FFT) or the Discrete Wavelet Transform (DWT).

The values quoted above are examples. Patient-to-patient variability is large and some need higher stimulation rates on the apical electrodes and/or lower stimulation rates on the basal electrodes. These are determined for each individual by evaluating a number of rate combinations in every day usage. Also, some patients do not have as many electrodes available and so the choice of electrodes is altered to suit their situation. However, the spectral ranges of the apical and basal electrodes remain much the same.

By using the DRSP program, features of speech will be more optimally presented to the cochlear implant user leading to improved speech understanding performance.

In the implementation of the second aspect of the invention, the software necessary to provide a variable rate of stimulation depending on the incoming speech signal is downloaded to the SPEAR processor and stored on an EPROM. Patient map details, including frequency bands, threshold (T) levels and comfort (C)

levels, are also stored on the device. Monopolar stimulation mode is used to reduce current levels and for longer battery life.

The standard rate of stimulation is 250 cycles/s and the higher rate is 1600 cycles/s. The patient's T and C levels are carefully set to ensure that stimulation levels are suitable for the two different rates. The electrodes to be stimulated are chosen by selecting the eight largest spectral energies within filterbanks derived from the Fast Fourier Transform (FFT) or the Discrete Wavelet Transform (DWT).

The changes in spectral energies and the amount of frequency energy are monitored over time. When there is a significantly large change between frames separated by the period of the lower stimulation rate then the higher stimulation rate is used for 50 ms. This procedure locates plosive bursts and other rapid spectral changes. The higher stimulation rate is also used when the ratio of energy below about 300Hz to that above about 2000Hz is less than about 0.5. This locates phonemes with significant frication.

The values quoted above are examples. Patient-to-patient variability is large and some need a higher stimulation rate for the standard rate and/or a lower stimulation rate for the higher rate. These are determined for each individual by evaluating a number of rate combinations in every day usage. Thresholds for changes in energy and ratio of energies are also adjustable for each individual.

The inventions described above have resulted from research undertaken by the inventors and described in the scientific paper annexed hereto as part of this specification.

Since other modifications within the spirit and scope of the invention may be readily effected by persons skilled in the art, it is to be understood that the invention is not limited to the particular embodiment described, by way of example, hereinabove.

DATED: 2 September 1999

CARTER SMITH & BEADLE
Patent Attorneys for the Applicants:

THE BIONIC EAR INSTITUTE

The Effect on Phoneme Recognition of the Rate of Stimulation of the Auditory Nerve

David B. Grayden[1] BE (Hons), BSc; Graeme M. Clark[1] FAA FTSE PhD;
Rodney D. Hollow[2] BSc, Dip. Aud.; and Katie L. Latus[2] BSp.Path., Dip. Aud.

[1]Human Communication Research Centre, The Bionic Ear Institute,
384-388 Albert St, East Melbourne VIC 3002, Australia

[2]Cochlear Implant Clinic, The Royal Melbourne Eye and Ear Hospital,
32 Gisborne St, East Melbourne VIC 3002, Australia

19 August, 1999

Corresponding Author:

Mr David B. Grayden

The Bionic Ear Institute

384-388 Albert Street

East Melbourne VIC 3002

AUSTRALIA

Telephone: +61 3 9283 7501

Facsimile: +61 3 9283 7518

Email: d.grayden@medoto.unimelb.edu.au

Corresponding Author

Mr David B. Grayden, BE (Hons), BSc

The Bionic Ear Institute

384-388 Albert Street

East Melbourne VIC 3002

AUSTRALIA

Telephone: +61 3 9283 7501

Facsimile: +61 3 9283 7518

Abstract

Objective: To determine the effect of increased stimulation rate on phoneme recognition for patients using a multiple-channel cochlear implant.

Design: Five adult patients received 24-consonant and 19-vowel syllable tests in quiet at three rates of stimulation: 250 pulses/s per channel, 807 pulses/s per channel, and 1615 pulses/s per channel. Eight channels selected from the maximum of up to 20 filterbanks were chosen for each presentation frame. The scores were analysed using a full factorial ANOVA model to compare overall results. The confusion patterns of consonants were examined using log-linear modeling and by ANOVA of a number of distinctive features to investigate the effect of rate of stimulation on the perception of the phonemes.

Results: There were no significant differences in phoneme recognition scores with increasing rate of stimulation. There was a high degree of patient variability, with each patient showing individual trends across the three rates. However, the errors that were made differed between the rates of stimulation. Analysis of perception of distinctive features showed that there tended to be fewer manner of articulation errors for the highest stimulation rate and fewer place of articulation errors for the lowest rate.

Conclusions: Increasing the rate of stimulation and processing does not necessarily increase patients' speech perception performance. There is a trade-off between the rates of stimulation for the perception of some distinctive features. This indicates the need to devise speech-processing strategies that provide a balanced rate for optimum performance, or adjust the rate of stimulation depending on the incoming speech.

INTRODUCTION

A key question for cochlear implants is the effect of rate of electrical stimulation on speech perception performance. This paper examines the effect of rate of stimulation on the perception of consonant and vowel phonemes. It extends the work performed by Vandali, Whitford, Plant, & Clark (1999) who studied the effect of rate of stimulation on the recognition of CNC words and of sentences in noise. They found that average speech recognition performance decreased with increased rate of stimulation. The present study extends this work by specifically examining confusion patterns between phonemes.

The perception of the waveform envelope and fine temporal structure is important for detection of voicing and manner of articulation (Van Tassel, Soli, Kirby, & Widin, 1987). Voicing is evidenced as periodicity in the speech waveform. Perception of manner of articulation depends on the amplitude envelope of the speech and the locations of transient events. For example, abrupt changes in energy at the onsets of closure and release are important temporal cues for stops, affricates, and nasals.

Further distinctions between phonemes are perceived using spectral characteristics that contain cues for the identification of place of articulation. These cues include formant locations and relative distributions of energy

throughout the spectrum. Duration is another important cue that can distinguish between fricatives and between vowels.

It is hypothesized that increasing the rate of stimulation of the auditory nerve will improve the ability of an implant user to correctly identify phonemes. A higher stimulation rate could improve the perception of voicing, manner of articulation, and changes in spectral characteristics by increasing temporal detail provided to the user. However, other influences could have a detrimental effect on the performance at high stimulation rates. For example, the impact of the refractory periods of the nerve fibers (Stypulkowski & van den Honert, 1984; Parkins, 1989; Bruce et al., 1999) on perception of electrical stimulation is not fully understood, especially for the complex electrical stimulation for speech sounds.

A prototype of the Advanced Combination Encoder (ACE) signal processing environment was used in this study with the CI24M implant. ACE is a flexible sound processing program, allowing the use of high rates of stimulation on up to 20 electrodes and various electrode selection techniques. A strategy similar to the SMSP strategy (McDermott, McKay, & Vandali, 1992) was used. The strategy commenced a cycle of stimulation by applying the fast Fourier transform (FFT) to the incoming sound. The frequency bands of the resulting digital spectrum were and then collapsed into up to 20 bins thus producing a reduced representation of the signal. The maximum

frequency range of the strategy extended from 80 to 7885 Hz. It employed filter bands that were linearly spaced up to 1125 Hz, and then logarithmically spaced up to 7885 Hz. The bins were assigned to electrodes in the implant in a tonotopic order. The lowest frequency bin was assigned to the most basal electrode available and the highest frequency bin was assigned to the most apical electrode. Eight bins with the highest amplitude were selected for each sampling period and electrodes associated with these bins were stimulated if their amplitude exceeded the threshold level of each electrode. Stimulation occurred in a basal-to-apical direction. This completed a stimulation cycle.

Patients were assessed in three conditions: 250 cycles/s (condition A), 807 cycles/s (condition B), and 1615 cycles/s (condition C). The analysis of the incoming speech data was performed at the same rate as the stimulation rates except for the 1615 cycles/s rate. For the highest rate, the speech information was sampled at about 800 Hz, but stimuli were presented twice to see how doubling the 807 cycles/s rate but providing no new information would affect speech perception. These quoted cycle rates are averages as 10% random timing jitter was introduced in order to reduce the effects of rate-pitch perception. Users of the cochlear implant perceive a pitch when stimulation occurs at a fixed rate (Tong, Miller, Clark, Martin, & Busby, 1980) and, when this is not related to the speech signal, it interferes with perception. The application of time jitter between pulses reduces the interaction of rate of

stimulation with perception of place-pitch and F0 modulation, especially at low rates of stimulation.

The inter-phase gap (the time between pulses of positive and negative polarity) used in conditions A and B was 25 μs. In condition C, the inter-phase gap was reduced to 8 μs in order to accommodate the desired stimulation rate. The opposite polarities were necessary to increase patient safety by ensuring that charges were balanced to remove direct current components.

Monopolar biphasic stimulation was used to reduce the current levels required to produce threshold (T) and comfortable (C) levels of perception compared to bipolar stimulation (Pfingst et al., 1995) and, thereby, increase battery life. This was at the expense of possibly increasing interaction between channels because of poorer current localization compared to bipolar stimulation methods (Millar, Tong, & Clark, 1984). However, speech perception has been shown to not suffer significantly as a result (Battmer, Martens, Gnadeberg, Hautle, & Lenarz, 1995) and is sometimes improved when using monopolar stimulation (Kileny, Zwolan, Boerst, & Telian, 1997).

METHOD

Subjects

Five patients implanted with the Nucleus CI24M receiver/stimulator manufactured by Cochlear Ltd participated in the experiment. All patients had been using the Spectral Peak (SPEAK) strategy (Seligman & McDermott, 1995) from two weeks post-operatively until the commencement of the study. The SPEAK strategy was based on the Spectral Maxima Sound Processor (SMSP) (McDermott, McKay, & Vandali, 1992). Speech was filtered using up to 20 band-pass filters and an average of six maximal outputs determined the sites and the amplitudes of stimulation. A rate of stimulation averaging 250 pulses/s per channel was used.

The patients were chosen based on availability and willingness to participate. Biographical data on the patients are presented in Table 1. The number of electrodes used by each patient varied between 15 and 20. The frequency range to electrode mapping was selected to provide similar frequency resolution between all five patients as shown in Table 2. The stimulus levels for threshold and maximum comfortable loudness on each active electrode were determined for each patient and rate condition.

Speech Material

The speech material were naturally produced consonants in /aCa/ context and naturally produced vowels in /hVd/ context. All of the 24 consonants and 19 vowels of Australian English were used. Lists of the consonants and vowels are presented in Tables 3 and 4, respectively.

The stimuli were recorded in an anechoic chamber by one male speaker (S1) and one female speaker (S2) who were both audiologists with extensive experience with live-voice testing. A studio-quality microphone was placed 0.5 m from the speaker in line with the forehead. The tokens were high-pass filtered at 70 Hz to eliminate room resonance and then sampled by a Pro-Audio Spectrum 16 Soundcard at 44.1 kHz with 16 bit samples. The utterances were normalised by placing 60 ms of silence before and after the syllables and a 20 ms ramp at each end to eliminate clicks. The tokens were then equalized to have the same RMS levels.

Two samples of each token were presented from CD-ROM using a soundcard through an amplifier and a single speaker in a sound-attenuated room. Each token was presented once in the randomised test sequences. The level at each

session was set to 75 dBA with the microphone of the sound level meter placed next to the ear of a subject sitting in the patients' chair.

Evaluation Procedure

Testing was performed using a repeated ABC protocol. The order of testing was balanced between the patients. Table 2 shows the order of testing that was used with each patient. The patients attended two testing sessions for each repetition of a strategy. The choice of starting with consonants or vowels was alternated between sessions and was randomised across the patients and across the strategies. Choice of male or female speaker to start with was also randomised across patients and was alternated between sessions.

In a testing session, the patients were first given a printed list of the phonemes from either the consonant or the vowel confusion sets. The phonemes were listed in alphabetical order in their syllable context as shown in the third columns of Tables 3 and 4. The syllables were then played to the patients, for the currently selected speaker, in the order displayed on their list. This was to allow the patients to familiarize themselves with the speaker.

After familiarisation was complete, two randomised lists of syllables uttered by the single speaker were presented to the patients. Each phoneme was

presented twice in each list. Responses were spoken by the patients and were recorded by clicking appropriate buttons on a screen using a mouse. Since the closed sets actually covered all possible consonant and vowel phonemes, the patients were encouraged to repeat exactly what they perceived.

After completion of the two lists, the syllables of the same type were presented using the other speaker's voice. The full syllable set was first played in alphabetical order to familiarize the patients with the new voice and then the two test lists were presented. The whole procedure was repeated for both speakers with the other class of phonemes.

RESULTS

The overall consonant and vowel recognition results for the five patients are shown in Figures 1 and 2, respectively. These figures show means and standard deviations when combining speakers and test-retest results.

The overall results were examined to see if there was an effect of rate of stimulation on consonant or vowel recognition performance. Analyses of

variance (ANOVA) were conducted on the total correct scores with factors: strategy, patient, speaker, and trial. The trial factor represented test-retest variability and accounted for learning effects. Careful attention was paid to significant interactions between factors since main effects must be viewed in the light of higher-order interactions that involve the same variable (Tabachnick & Fidell, 1996). Where the effect of patient or strategy was significant, Tukey's procedure ($p < 0.05$) was used to make comparisons between pairs (Devore, 1987).

Consonants

Analyses of variance conducted on the consonant scores showed all main effects were significant. The largest effect was between patients ($F[4,240] = 247.6$, $p < 0.001$). Tukey's procedure showed that all but two of the patients had significantly different results from each other. The differences between the patients' results can be seen in Figure 1.

The average effect of strategy was significant ($F[2,240] = 9.7$, $p < 0.001$). Tukey's test showed that there was no significant difference between the 250 cycles/s and 807 cycles/s rates of stimulation. However, the average performances of both of these strategies were significantly better than the 1615 cycles/s rate of stimulation.

Overall, the average performance with speaker 1 was greater than speaker 2 ($F[1,240] = 44.1$, $p < 0.001$). Trial 2 produced a better score than trial 1 ($F[1,240] = 18.0$, $p < 0.001$) which shows that there was a significant learning effect.

In summary, there was a significant difference between performances with the different rates of stimulation when the test-retest scores were averaged together. There was a significant difference between the overall recognition abilities of the patients showing wide variance in the patients' results. There was also a difference between performances with the two speakers and a strong learning effect as the patients familiarized themselves with the different strategies over the duration of the study.

However, the interpretation of the results stated above must account for interactions between the factors. Each significant interaction shows that a particular factor was influenced by another factor. The significant interactions for consonant recognition performance were strategy-patient ($F[8,240] = 4.6$; $p < 0.001$), patient-speaker ($F[4,240] = 8.8$; $p < 0.001$), strategy-patient-speaker ($F[8,240] = 5.0$; $p < 0.001$), patient-trial ($F[4,240] = 7.5$; $p < 0.001$), and strategy-patient-trial ($F[8,240] = 2.6$; $p = 0.010$). Interpretations of these interactions are described below.

The strategy-patient interaction is obvious in Figure 1. Patients 1, 2 and 4 followed the same trend as the overall average, although patients 1 and 4 had steeper declines with increasing rate of stimulation. Patients 3 and 5 had different trends, with the performance of strategy B tending to be less than strategies A and C.

There is also a patient-speaker interaction, which indicates that the preferred speaker differed between patients. Figure 3 illustrates the differences between the patients' scores with different speakers for all strategies combined. It is clear that there was a trend of better performance with speaker 1 than speaker 2, but there is a significant difference between the patients on the size of this trend.

The patient-strategy-speaker interaction shows either that the interaction between patient and strategy varied with speaker or that the interaction between patient and speaker varied with strategy, or both.

The patient-trial interaction indicated that the there were different learning effects between patients. The patient-strategy-trial interaction shows that individual patient learning rates differed depending on the strategy. The latter interaction is of concern since it results from greater improvement between over the duration of the study for the higher-rate strategies than for the lower-rate strategy. This was explored by performing ANOVA using only data from

the second trial. The factors were patient, strategy and speaker. The significant main effects were patient ($F[4,120] = 164.2$; $p < 0.001$) and speaker ($F[1,120] = 17.7$; $p < 0.001$). Strategy was no longer significant ($F[2,240] = 1.2$; $p = 0.309$). This shows that there were no overall significant differences between the three rates of stimulation when the patients were given sufficient time to adjust to all of the strategies.

Results using data from the second trial maintained the patient-strategy ($F[8,120] = 2.3$; $p = 0.030$) and patient-speaker ($F[4,120] = 11.0$; $p < 0.001$) interactions. However, the patient-strategy-speaker interaction was no longer significant ($F[8,120] = 1.5$; $p = 0.160$) because of decreased differences between strategies in the second trial.

The patient-strategy interaction was explored by examining the results for patients separately. ANOVA were performed for each patient using the factors strategy and speaker for data obtained in trial 2. Strategy was only a significant main effect for patient 1 ($F[2,24] = 5.6$; $p = 0.013$). Tukey's test showed that the significant difference between strategies for this patient was that the performance was better for strategy B than strategy C. Strategy A was not significantly different from either of the other two strategies.

Patients 2 and 4 showed significantly better performance with speaker 1 than speaker 2. However, patient 3 performed better with speaker 2 than speaker 1. There were no significant effects for patient 5.

In summary, when only considering the second trial, there was no significant overall effect of rate of stimulation on consonant recognition performance. However, there were large differences between patients' scores and performances with the different speakers.

Vowels

ANOVA were also performed for the vowel scores. As for the consonants, all main effects were significant. There were very large differences between patients ($F[4,240] = 288.8$; $p < 0.001$) and between speakers ($F[1,240] = 32.7$; $p < 0.001$). The average strategy effect was significant ($F[2,240] = 7.3$; $p = 0.001$) with Tukey's test showing that strategy A (250 cycles/s) was better than strategy C (1615 cycles/s). Strategy B (807 cycles/s) was not significantly different from either of the other two strategies.

There were significant interactions between strategy and patient ($F[8,240] = 2.3$; $p = 0.025$) and between patient and speaker ($F[4,240] = 31.5$; $p < 0.001$). As for the consonants, there was a learning effect ($F[1,240] = 29.5$; $p < 0.001$)

that interacted with strategy ($F[2,240] = 3.7$; $p = 0.027$). There were also strategy-patient-trial ($F[8,240] = 2.105$; $p = 0.037$) and patient-speaker-trial ($F[4,240] = 4.7$; $p = 0.001$) interactions.

When only the second trial was considered, ANOVA with factors patient, strategy and speaker produced no overall main strategy effect ($F[2,120] = 1.0$; $p = 0.355$). The significant effects were patient ($F[4,120] = 178.7$; $p < 0.001$) and speaker ($F[1,120] = 12.1$; $p = 0.001$) with interactions strategy-patient ($F[8,120] = 2.1$; $p = 0.040$) and patient-speaker ($F[4,120] = 7.4$; $p < 0.001$).

When patients were analysed separately, none showed a significant strategy effect when only considering the second trial. However, significant differences in speakers were maintained. Results were significantly better for speaker 2 than for speaker 1 for patients 1, 4 and 5. Speaker 1 was better than speaker 2 with patient 3. There was no speaker effect for patient 2.

In summary, there was no significant effect of rate of stimulation on vowel recognition performance when only considering the second trial. However, there were large differences between patients and speakers.

Log-Linear Modeling

The consonant confusion matrices were analysed using hierarchical loglinear modeling to see if there was an effect of rate of stimulation on the pattern of phoneme confusions. The analyses of variance performed above only examined the effect of rate of stimulation and other factors on the number of correct responses without paying attention to the types of errors that were made. Hierarchical loglinear modeling takes this a step further by determining the significant factors and interactions that are necessary to describe the patterns of responses; in this case, to see if responses varied with the different rates of stimulation. Only the consonants were considered because vowel recognition performance was high resulting in very sparse confusion matrices.

Hierarchical loglinear analysis allows the observed values in a confusion matrix to be compared with predicted values obtained from the products of marginal frequencies by using the likelihood-ratio chi-square approximation (Bell, Dirks, Levitt, & Dubno, 1986). The goal is to determine a minimal number of factors and interaction terms that are required to predict the observed confusion matrix thus allowing easier interpretation of the data. Using a process of backward elimination (Bishop, Fienberg, & Holland, 1975), interaction terms and factors that are not significant can be successively discounted from the model. The final model will show the factors and interactions that are necessary to adequately describe the confusion matrix.

As in the consonant confusion work of Bell et al. (1986), only the error frequencies were considered. The correct responses were removed by considering the diagonal entries of the confusion matrices to be structural zeros (Bishop, Fienberg, & Holland, 1975). Only data from the second trial was used because this better represented the performances for the different strategies. The results for each patient and speaker combination were examined separately because of the diversity of the results. Thus, the factors were strategy, stimulus, and response. There were three levels of strategy, each representing one of the rates of stimulation. Stimulus represented the phonemes presented to the patients. Each level of this factor was a different phoneme so there were 24 levels in all for the consonants. The response factor represented the responses that the patients gave. Again, there were 24 levels for this factor that represented the perceived consonant phonemes.

For each patient-speaker combination, phonemes that were either never wrongly perceived or were never an error response were excluded from the confusion matrices. This was necessary since they were rows or columns with zero entries in the confusion matrices when the matrix diagonals (correct responses) were not considered and so could not be handled by the loglinear procedure.

The basis of hierarchical loglinear modeling is to find a minimal model that can sufficiently represent the data by successively removing high-order

interaction terms. Thus, with three factors, the first step was to determine if the three-way interaction term, strategy-stimulus-response, was required. In all cases, its contribution was not significant. Therefore, it did not need to be considered and the model could be simplified. Then the contribution of each two-way interaction term was investigated. These terms were strategy-stimulus, strategy-response, and stimulus-response.

The models that were derived for each patient-speaker combination are shown in Table 5. Since the models were hierarchical, lower-order terms are not shown if they appear in a higher-order interaction. When there was no strategy interaction term, the influence of strategy as a main effect was also tested.

In all cases there were highly significant interactions between stimulus and response. This was expected because there are varying degrees of differences between the phonemes. All of the patients displayed an ability to distinguish between the more dissimilar phonemes. For example, there were no confusions between /μ/ and /σ/ because the properties of these phonemes are so different.

The significance of strategy and of the remaining two interaction terms, strategy-stimulus and strategy-response, varied between patients. Patient 1 with speaker 1 and patient 2 with either speaker showed no strategy effect.

Their patterns of confusions did not vary significantly with different rates of stimulation.

The other patient-speaker combinations all exhibited significant interactions between strategy and response. Thus, the patterns of responses for the consonant phonemes varied significantly with rate of stimulation. However, for all but one case there was no significant interaction between strategy and stimulus. This shows that approximately the same number of errors was made for each phoneme across the three strategies but the types of errors made for each phoneme varied with different rates of stimulation.

Patient 4 with speaker 1 also required a strategy-stimulus term. In this case, some phonemes were better perceived with a particular rate of stimulation. However, since the overall number of correct responses did not vary significantly between strategies, as shown by ANOVA above, then where there was an improvement for some phonemes at a particular rate of stimulation there was also an equivalent reduction in performance for other phonemes.

Analysis of Distinctive Features

In order to investigate the relationships between rate of stimulation and error responses, the perceptions of a number of distinctive features were analysed. Binary distinctive features were chosen from those described by Miller &

Nicely (1955), Singh (1968), Chomsky & Halle (1968) and Singh, Woods, & Becker (1972). Most of these are summarized by Singh (1976) and Edwards (1992). The presentations of the two aspects of each feature were considered to be independent. For example, the number of unvoiced phonemes presented, and their resulting perceptions as voiced or unvoiced, was considered independent of the number of voiced phonemes presented. Thus, the number of correct responses for each aspect of each feature could be analysed by separate analyses of variance. Careful attention was paid to the interactions between variables so that only significant results were reported. Tukey's procedure ($\underline{p} < 0.05$) was used to compare strategies when a significant effect was discovered.

A list of the distinctive features that were investigated is provided in Table 6. Several binary features, such as vocalic, rounded, low, and lateral were not included since they separate one or two phonemes from the remainder. The feature tense was also not used since it is very similar to voicing. The vocalic consonants (/λ, ρ/) are those that are produced with little vocal tract constriction and so are similar to vowels. Rounded sounds are made with pursed lips and only include the consonants /ρ/ and /ω/ in English. There is only one low consonant, /h/, which is produced with the body of the tongue lowered below the neutral position for /↔/. Lateral sounds are produced with the airflow passing around the sides of the tongue and, in English, only

include the phoneme /λ/. Tense consonants are produced with increased muscular contraction at the tongue root compared to lax sounds. Except for the phonemes /δZ/, /η/, and /λ/, +tense consonants are –voiced and –tense phonemes are +voiced.

The number of correct responses for each aspect of each feature was examined to see if strategy influenced the perception of the feature. Only results from the second trial were used. The analysis procedure was to first perform a three-way ANOVA with factors strategy, patient, and speaker. If there were no interactions between strategy and either of the other two factors, then the analysis stopped at this point. However, if there was an interaction involving strategy then further analyses were performed to investigate the impact of the interaction on individual patient or speaker results by splitting the data. If an interaction was found between strategy and the remaining factor in the second analysis, one-way ANOVA were performed for individual patient-speaker combinations. When the effect of strategy was significant ($p < 0.05$), Tukey's test was applied with $p < 0.05$ to determine the ordering of strategy performance.

The factors patient and speaker and the patient-speaker interaction were significant in most cases. This continued to show the variability in the overall performance levels of the patients and their varying abilities with the two speakers.

Manner of Articulation Features

The analyses of the manner of articulation features shown in Table 6 will be described first. There were no significant strategy main effects for +sonorant or -sonorant, nor were there any interactions between strategy and either of the other factors. This means that the perception of these distinctive features was not significantly affected by different rates of stimulation. Sonorant phonemes are those produced without significant obstruction of the vocal tract, i.e., the nasals, liquids and glides.

There was a significant strategy effect on the perception of +nasal ($F[2,120] = 3.129$, $p = 0.049$) and no significant interactions between strategy and the other factors. There was a trend of C-A-B going from best to worst, but Tukey's test, which is more stringent, showed that there were no significant differences between the rates of stimulation ($p = 0.068$). When individual patient's results were analysed separately, it was found that for patient 4, strategy C was significantly better than strategy A. The feature –nasal showed no significant strategy effects.

Continuant phonemes are those produced without a complete constriction anywhere within the vocal tract. The onset and offset of a constriction creates a discontinuity in the envelope of energy for interrupted phonemes as

observed for plosives, affricates, and nasals (Jakobson, Fant, & Halle, 1951). The feature +continuant did not have a significant main strategy effect, but there was an interaction between strategy and patient. Separating the patients showed that only patient 3 had a significant strategy effect ($F[2,24] = 6.264$, $p = 0.009$) and no significant interaction with speaker. Tukey's test revealed that strategy C was significantly better for correct perception of +continuant than strategies A and B for patient 3. There was an overall significant effect of strategy on the perception of -continuant ($F[2,120] = 4.576$, $p = 0.013$) and no significant strategy interactions. Tukey's test revealed that strategy C was significantly better than strategy B for the correct perception of –continuant but neither was significantly different from strategy A.

The feature +voiced showed no significant strategy effects or interactions with strategy. However, strategy had a significant effect on the correct perception of the feature -voiced ($F[2,120] = 3.258$, $p = 0.043$) and no significant interactions. Tukey's test showed that strategy C was significantly better than strategy A.

The three features frication, strident, and sibilant respectively include fewer of the fricatives in their positive aspect. Frication includes all fricatives and affricates while strident only includes strong fricatives (/δZ, τΣ, ζ, Z, σ, Σ, φ, ϖ/), leaving out the weaker ones (/T/, /Δ/, and /η/). Sibilant phonemes (/δZ, τΣ, ζ, Z, σ, Σ/) are produced by directing airflow

against a hard surface such as the hard palate and the teeth producing considerable noise. There were no significant effects or interactions involving strategy for +frication, -frication, +strident, -strident, or +sibilant. However, the effect of strategy on the correct perception of –sibilant was significant ($F[2,120] = 5.439$, $p = 0.006$). For this feature, Tukey's test showed that strategy C was significantly better than strategy A for perceiving –sibilant phonemes.

Duration further reduces the number of fricatives included in sibilant as it only includes the longer fricatives /σ/, /ς/, /Σ/, and /Z/. The feature +duration had no main strategy effect but significant strategy-patient and strategy-patient-speaker interactions. Further analyses showed that strategy was significant for patient 1 ($F[2,24] = 5.744$, $p = 0.012$) with Tukey's test showing that strategy C was significantly worse than strategies A and B. Strategy was also significant for patient 5 with speaker 2 ($F[2,12] = 5.087$, $p = 0.033$) and strategy C was significantly better than strategy A. Strategy significantly affected the correct perception of –duration ($F[2,120] = 4.164$, $p = 0.019$) and there were no strategy interactions. Tukey's test showed that strategy A was significantly worse than strategies B and C.

Place of Articulation Features

The place of articulation features are also shown in Table 6. Anterior phonemes are produced with the point of constriction anterior to /Σ/. They were called diffuse by Jakobson, Fant, & Halle (1951) as opposed to compact for –anterior phonemes. These terms relate to spectral characteristics. Compact phonemes have strong concentrations of energy in the mid-frequency region while diffuse phonemes have energy concentrations at low or high frequencies. There were no significant strategy main effects or interactions between strategy and either of the other factors for +anterior. However, there was a main strategy effect for -anterior ($F[2,120] = 3.166$, $p = 0.047$) but this was accompanied by significant strategy-patient and strategy-speaker interactions. Further analyses showed that strategy was significant for patient 5 ($F[2,24] = 7.915$, $p = 0.003$) with strategy A significantly better than strategies B and C. Strategy significantly affected the results for speaker 1 with all patients ($F[2,60] = 3.766$, $p = 0.031$) and strategy A was significantly better than strategy B. Patient 1 with speaker 2 showed a significant strategy affect ($F[2,12] = 17.271$, $p = 0.001$) that differed from the above results as strategy A and C were significantly worse than strategy B.

Coronal phonemes are produced with the tongue blade raised above the neutral position. The neutral position is that used to produce the vowel /↔/. Jakobson, Fant, & Halle (1951) separated diffuse phonemes in this way by calling them acute or grave for +coronal and –coronal, respectively. Acute phonemes have energy concentrated at high frequencies while grave

phonemes have energy concentrated at low frequencies. There was a main effect of strategy for +coronal ($F[2,120] = 3.943$, $p = 0.023$) but with significant strategy-patient-speaker interaction. However, when the data was split by patient, speaker, or both, the significant main effect of strategy was no longer seen. The feature –coronal had no significant strategy effect or interaction.

High phonemes are produced with the tongue body raised above the neutral position. These are the palatal and back consonants as well as the glides. The feature +high had a significant main effect of strategy ($F[2,120] = 3.294$, $p = 0.042$) and a significant interaction between strategy and speaker. Separating the speakers revealed that strategy significantly affected each speaker's results (S1: $F[2,60] = 3.897$, $p = 0.027$; S2: $F[2,60] = 4.692$, $p = 0.014$). However, the ordering of the strategies for the speakers was different as revealed by Tukey's test. For speaker 1, strategy A was better than strategy B while neither differed significantly from strategy C. For speaker 2, strategies A and B were both significantly better than strategy C. There were no significant effects of interactions of strategy on correct perception of –high.

Back phonemes are produced by retracting the tongue to the back of the mouth. The feature +back was significantly influenced by strategy ($F[2,120] = 3.628$, $p = 0.031$) while not having any significant strategy interactions. Strategy A was better for correctly perceiving this feature than strategy B.

The feature –back did not have a significant strategy main effect but had significant strategy-patient and strategy-patient-speaker interactions. Further analysis showed that patient 5 had a significant strategy effect ($F[2,24] = 3.664$, $p = 0.046$) with strategy C better than strategy A. Similarly, patient 4 with speaker 1 had a significant strategy effect ($F[2,12] = 8.029$, $p = 0.010$) with strategy C better than strategy A.

Distributed phonemes are produced with a relatively long constriction along the direction of airflow. There was no main effect of strategy on +distributed but there were strategy-patient, strategy-speaker, and strategy-patient-speaker interactions. Separating the patients revealed that strategy was significant for patient 5 ($F[2,24] = 4.233$, $p = 0.031$) with strategy A better than strategy C. Separating the speakers showed that strategy was significant for speaker 1 over all patients ($F[2,60] = 5.642$, $p = 0.007$) with strategy A also better than strategy C. The three-way interaction strategy-patient-speaker was manifest in a significant effect of strategy on patient 1 with speaker 2 ($F[2,12] = 25.800$, $p < 0.001$) for whom strategy B was significantly better than strategies A and C. There was no effect of strategy for –distributed.

The significant results for the distinctive features are summarized in Table 7.

DISCUSSION

There was no significant difference in phoneme recognition performance for the different rates of stimulation once learning effects were taken into account. However, the rate of stimulation affected the perception of some phonetic distinctive features. There was a trade-off between the strategies in the recognition of features that resulted in no change in performance overall.

Examination of Table 7 shows that there was a tendency for manner of articulation features to be better perceived with the higher rates of stimulation. Over all patients and speakers, strategy C was significantly better than strategy A for the perception of –voiced, –sibilant, and –duration. The features +nasal, +continuant, and +duration also showed this effect for particular patients. The –continuant feature also showed this trend, but not significantly.

The only exception was the perception of +duration by patient 1. This was probably caused by the overall difficulty that this patient had in adjusting to strategy C. She was the only one to have a significant difference in phoneme recognition performance, with strategy C being significantly worse than strategy B.

These trends were further examined by performing information transmission analyses (Miller & Nicely, 1955) for these features. This was done by collapsing the confusion matrices down to 2X2 matrices for each feature and each strategy. Then the percentage of information transmitted was calculated

for each reduced matrix. The results are shown in Table 8 and Figure 4. The percentage of information transmitted using strategy C was always greater than that transmitted using strategy A. Strategy C was also better in transmitting information about nasality and continuancy than strategy B. Strategy B had greater information transmission than strategy C for voicing, sibilancy, and duration.

Place of articulation features were better recognised with the lower rates of stimulation. Strategy A was generally better than the higher rate strategies for the perception of –anterior, +high, +back, and +distributed features. There are a number of exceptions to these observations that were caused by patient and speaker variability. There are again some variations for patient 1 which result from the greater differences between strategies observed in her results. Also, strategy C was significantly better for the perception of –back for patient 4 with speaker 1 and for patient 5. However, the other three patients showed trends towards C being the worst strategy for this feature.

These place of articulation observations were confirmed by further information transmission analyses shown in Table 8. In all cases, strategy C was the poorest for transmitting the place of articulation information. There was little difference between strategies A and B for these tasks.

Therefore, none of the different rates of stimulation were the best for conveying all information about phoneme identity. The higher rates of stimulation were better able to convey envelope information as evidenced by improved perception of nasality, continuancy, voicing, and duration. The higher rates of stimulation were also better able to portray high frequency noise that helps to distinguish the sibilants from other phonemes.

However, the lower rates of stimulation were better for presenting place of articulation information to the users. Distinguishing these features requires perception of spectral detail such as formant locations and general spectral distributions. This highest rate of stimulation appears to have masked these fine spectral characteristics. The rate of processing for strategy C was not different to strategy B, but the perception of spectral structure was much better for the latter strategy. The greater pulse rate, while able to convey good temporal information, was making the perception of place of stimulation more difficult.

The problem with the higher rates of stimulation can be explained with reference to the refractory effects of auditory nerve (AN) fibers (Stypulkowski & van den Honert, 1984; Parkins, 1989). After generation of an action potential there is an absolute refractory period, lasting for about 0.7ms for the cat AN (Bruce et al., 1999), during which discharge is impossible. This is followed by the relative refractory period when increased current is required to

cause the neuron to discharge. The relative refractory period lasts for up to 20ms during which the required threshold decreases exponentially (Bruce et al., 1999) with the major influence of the effect reducing significantly during the first 4ms (Stypulkowski & van den Honert, 1984).

The inter-stimulus intervals for stimulus pulses with the three rates of stimulation were, on average, 4ms for the 250 cycles/s rate, about 1.2ms for the 807 cycles/s rate, and approximately 0.6ms for the 1615 cycles/s rate. The time between pulses for the 250 cycles/s rate is just outside the effective relative refractory period so there should be little effect at this rate of stimulation. However, the 807 cycles/s rate is within the relative refractory period and the 1615 cycles/s rate is close to the absolute refractory period of the auditory nerve. This would have a significant effect on the perception of these stimuli by the patients. For the higher-rate strategies, if an electrical pulse delivered at a particular site is close to comfort (C) level, then a large number of neurons close to that site will fire. The next pulse to be delivered at the same site will occur during the refractory period and so will produce much less discharge across the bundle of neurons. The implication of this is that amplitude will not be correctly transmitted from the cochlea to the higher level processes. The effect of current spread will add to this problem. Place of stimulation is important for perception of spectral detail. It is blurred somewhat by current spread which causes neurons to discharge some distance from the stimulating electrode. If the nerve fibers nearest to the stimulating

electrode are suffering from refractory effects, then it is possible that there will be more action potential generated by the surrounding neurons that fire due to current spread, than the nearest neurons. This will blur the spectral detail provided to the patient leading to increased difficulty in perceiving place of articulation cues. However, the current spread will help to ensure that timing information will be preserved since action potentials are still being generated and so not will not affect the perception of manner of articulation cues.

A very similar effect of blurring of spectral detail was discovered by ter Keurs, Festen, & Plomp (1991). They tested consonant and vowel perception with normal-hearing listeners when speech was smeared over bandwidths of up to 2 octaves in the frequency regions from 100 to 8000 Hz. At the 2-octave level of smearing, the place of articulation of consonants became confused while manner of articulation was still well perceived. Spectral smearing is a problem for cochlear implant users because of the small number of electrodes that are used. The effect of current spread exacerbates this problem.

The difference in perception between rates of stimulation gives insight into why the information transmission for voiced, sibilant and duration was best with strategy B. These features require the use of spectral information as well as temporal information. Voicing is not only represented by periodicity in the waveform and by voice onset time for plosives, but also by the voice bar and its harmonics in narrow-band spectra. Information about frication is conveyed

by rapid fluctuations in the waveform and also by the balance between low and high frequency energy. The intermediate stimulation rate struck a balance between these envelope and spectral cues and so gave good performance.

CONCLUSIONS

Different rates of electrical stimulation provide different benefits for consonant recognition. Higher rates of stimulation give improved manner of articulation information compared to low rates because increased temporal resolution provides more information about variations in amplitude envelope. Higher rates also improve perception of some fricative sounds by better representing noise. However, perception of place of articulation cues is reduced with high stimulation rates, probably because of smearing of spectral information due to refractory periods of the auditory neurons. The correct perception of phonemes requires perception of both manner of articulation and place of articulation so there was no overall improvement in recognition between the three rates of stimulation that were investigated.

ACKNOWLEDGEMENTS

This work was supported by the Sidney Myer Fund and the Bionic Ear Institute. We are grateful to the participants who were involved. We also thank Mr. Andrew Vandali, Ms. Kerry Plant and Ms. Lesley Whitford from the Cooperative Research Centre for Cochlear Implant, Speech and Hearing Research for their help with the research participants. We thank Dr. Chris James for recording and editing the speech stimuli and writing the testing program, Dr. Richard Dowell and Ms. Karyn Galvin for uttering the speech samples, and Prof. Ray Watson for his help with statistical analyses.

REFERENCES

Battmer, R. -D., Martens, U., Gnadeberg, M. S., Hautle, M. S., & Lenarz, T. (1995). Comparison study of patients using either the Nucleus minisystem-22 in bipolar mode or the Nucleus 20 + 2 in monopolar mode. Annals of Otology, Rhinology, and Laryngology, Supp. 166, 349-351.

Bell, T. S., Dirks, D. D., Levitt, H., & Dubno, J. R. (1986). Log-linear modeling of consonant confusion data. Journal of the Acoustical Society of America, 79, 518-525.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: The MIT Press.

Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., Dynes, S., Javel, E., & Clark, G. M. (1999). A stochastic model of the electrically stimulated auditory nerve: pulse-train response. IEEE Transactions on Biomedical Engineering, 46, 630-637.

Chomsky, N. & Halle, M. (1968). The Sound Pattern of English. New York: Harper & Row.

Devore, J. L. (1987). Probability and Statistics for Engineering and the Sciences. Belmont, CA: Wadsworth, Inc.

Edwards, H. T. (1992). Applied Phonetics: The Sounds of American English. San Diego: Singular Publishing Group, Inc.

Jakobson, R., Fant, G. & Halle, M. (1951). Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates. Cambridge, MA: MIT Press.

Kileny, P. R., Zwolan, T. A., Boerst, A., & Telian, S. A. (1997). Monopolar vs bipolar stimulation: Psychophysical and speech recognition results. 1997 Conference on Implantable Auditory Prostheses, Asilomar Conference Center, Pacific Grove, California, August 17-21, 38.

McDermott, H. J., McKay, C. M., & Vandali, A. E. (1992). A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. Journal of the Acoustical Society of America, 91, 3367-3371.

Millar, J. B., Tong, Y. C., & Clark, G. M. (1984). Speech processing for cochlear implant prostheses. Journal of Speech and Hearing Research, 27, 280-296.

Miller, G. A. & Nicely, P. E. (1955). An analysis of perceptual confusions among English consonants. . Journal of the Acoustical Society of America, 27, 338-352.

Parkins, C. W. (1989). Temporal response patterns of auditory nerve fibers to electrical stimulation in deafened squirrel monkeys. Hearing Research, 41, 137-168.

Pfingst, B. E., Miller, A. L., Morris, D. J., Zwolan, T. A., Spelman, F. A., Clopton, B. M. (1995). Effects of electrical current configuration on stimulus detection. Annals of Otology, Rhinology, and Laryngology, Supp. 166, 127-131.

Seligman, P. S. & McDermott, H. J. (1995). Architecture of the Spectra 22 speech processor. Annals of Otology, Rhinology, and Laryngology, Supp. 166, 139-141.

Singh, S. (1968). A distinctive feature analysis of responses to a multiple choice intelligibility test. International Review of Applied Linguistics, 6, 37-53.

Singh, S. (1976). Distinctive Features: Theory and Validation. Baltimore: University Park Press.

Singh, S., Woods, D. R. & Becker, G. M. (1982). Perceptual structure of 22 prevocalic English consonants. Journal of the Acoustical Society of America, 52, 1698-1713.

Stypulkowski, P. H. & van den Honert, C. (1984). Physiological properties of the electrically stimulated auditory nerve. I. Compound action potential recordings. Hearing Research, 14, 205-223.

Tabachnick, B. G. & Fidell, L. S. (1996). Using Multivariate Statistics. New York, NY: HarperCollins College Publishers.

ter Keurs, M., Festen, J. M., & Plomp, R. (1991). Effect of spectral envelope smearing on speech reception. I. Journal of the Acoustical Society of America, 91, 2872-2880.

Tong, Y. C., Miller, J. B., Clark, G. M., Martin, L. F., & Busby, P. A. (1980). Psychophysical studies for a multiple-channel cochlea implant. 10th International Congress on Acoustics, Sydney, B12.4.

Van Tassel, D., Soli, S., Kirby, V., & Widin, G. (1987). Speech wave-form envelope cues for consonant recognition. Journal of the Acoustical Society of America, 82, 1152-1161.

Vandali, A. E., Whitford, L. A., Plant, K. L., & Clark, G. M. (1999). Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system. Manuscript submitted for publication.

TABLE 1: Biographical data on the implant patients.

Patient	Age at Testing	Months Implanted	Duration of severe-to-profound deafness (years)	Etiology of Hearing Loss
P1	63	8	0.75	Unknown
P2	68	10	0.75	Unknown
P3	44	6	< 10	Otosclerosis
P4	70	5	5	Chronic otitis media
P5	46	18	0.5	Unknown/progressive

TABLE 2: Map and evaluation details: number of electrodes active in speech processor map, frequency range of speech processor map, and order of evaluation of the three rate conditions. The conditions are 250 cycles/s (A), 807 cycles/s (B), and 1615 cycles/s (C).

Patient	Number of electrodes in map	Frequency range (Hz)	Rate condition evaluation order
P1	16	160-5744	ABC ABC
P2	20	116-7871	CAB CAB
P3	15	244-4177	BCA BCA
P4	20	116-7871	CBA CBA
P5	18	142-7009	BAC BAC

TABLE 3: Consonant phoneme set. Twenty-four consonants were uttered in /aCa/ context. The first column shows the phonemes investigated and the second column shows the context in which they were uttered. The third column shows how the phoneme was written for the patients on the sheet that they were given at each session. The phonemes are listed here by manner of articulation. The written list was given to patients in alphabetical order.

Phoneme	Context	Written	Phoneme	Context	Written
/β/	/αβα/	ABA	/σ/	/ασα/	ASA
/δ/	/αδα/	ADA	/Σ/	/αΣα/	ASHA
/γ/	/αγα/	AGA	/ϕ/	/αϕα/	AFA
/π/	/απα/	APA	/T/	/αTα/	ATHA
/τ/	/ατα/	ATA	/η/	/αηα/	AHA
/κ/	/ακα/	AKA	/μ/	/αμα/	AMA
/δZ/	/αδZα/	AJA	/ν/	/ανα/	ANA
/τΣ/	/ατΣα/	ACHA	/N/	/αNα/	ANGA
/ζ/	/αζα/	AZA	/λ/	/αλα/	ALA
/Z/	/αZα/	AZHA	/ρ/	/αρα/	ARA
/ϖ/	/αϖα/	AVA	/ω/	/αωα/	AWA
/Δ/	/αΔα/	ATTHA	/φ/	/αφα/	AYA

TABLE 4: Vowel phoneme set. Nineteen vowels were uttered in /hVd/ context. The first column shows the phonemes investigated and the second column shows the context in which they were uttered. The third column shows how the phoneme was written for the patients on the sheet that they were given at each session. The phonemes have been listed here in alphabetical order as listed for the patients.

Phoneme	Frame	Written	Phoneme	Frame	Written
/Θ/	/ηΘδ/	HAD	/αI/	/ηαIδ/	HIDE
/εI/	/ηεIδ/	HADE	/ /	/η δ/	HOARD
/E↔/	/ηE↔δ/	HAIRD	/ /	/η δ/	HOD
/α/	/ηαδ/	HARD	/oY/	/ηoYδ/	HODE
/↔/	/η↔δ/	H'D	/ I/	/η Iδ/	HOID
/E/	/ηEd/	HEAD	/Y/	/ηYδ/	HOOD
/ι/	/ηιδ/	HEED	/αY/	/ηαYδ/	HOW'D
/∈/	/η∈δ/	HERD	/℘/	/η℘δ/	HUD
/I↔/	/ηI↔δ/	HERE'D	/υ/	/ηυδ/	WHO'D
/I/	/ηIδ/	HID			

TABLE 5: Consonant hierarchical log-linear models for patient-speaker pairs. Only the data obtained from the second trial was used. The factors are strategy (SY), stimulus (SS) and response (RE). Since the models were hierarchical, listing an interaction means that the lower order terms were also included in the model.

	Hierarchical log-linear model	
Patient	Speaker 1 (male)	Speaker 2 (female)
P1	SS*RE	SY*RE + SS*RE
P2	SS*RE	SS*RE
P3	SY*RE + SS*RE	SY*RE + SS*RE
P4	SY*SS + SY*RE + SS*RE	SY*RE + SS*RE
P5	SY*RE + SS*RE	SY*RE + SS*RE

TABLE 6: Phonetic distinctive features and their definitions (Edwards, 1992, Singh, 1976). These definitions describe the positive state of the features. The negative state is the opposite.

Feature	Definition
Manner of articulation features	
Sonorant	Relatively open vocal tract that allows resonation.
Nasal	The oral tract is closed and air flows through the nose.
Continuant	Airflow is not blocked at any point in the vocal tract.
Voiced	Vibration of the vocal folds.
Frication	Air forced through a narrow aperture creating noise.
Strident	Considerable noise is produced.
Sibilant	Considerable high-frequency noise is produced.
Place of articulation features	
Anterior	Obstruction anterior to location for /Σ/.
Coronal	Tongue blade raised above neutral position.
High	Tongue body raised above neutral position.
Back	Tongue retracted to back of mouth.
Distributed	Constriction extends for a relatively long distance along the vocal tract.

TABLE 7: Significant differences between strategies for various aspects of phonetic distinctive features. Where a particular patient (e.g., P4) or speaker (e.g., S1) is given, that effect was subject to interactions and so was only reported for that particular patient, speaker, or combination of the two. The conditions are 250 cycles/s (A), 807 cycles/s (B), and 1615 cycles/s (C).

Feature	Differences between strategies
Manner of articulation features	
+nasal	P4: C > A
+continuant	P3: C > AB
-continuant	C > B
-voiced	C > A
-sibilant	C > A
+duration	P1: AB > C; P5S2: C > A
-duration	BC > A
Place of articulation features	
-anterior	P5: A > BC; S1: A > B; P1S2: B > AC
+high	S1: A > B; S2: AB > C
+back	A > B
-back	P5: C > A; P4S1: C > A
+distributed	P5: A > C; S1: A > C; P1S2: B > AC

TABLE 8: Percentage of information transmitted for phoneme distinctive features that produced significant differences between strategies. The information transmission was computed for each strategy using combined data from all five patients and both speakers but only used results from the second trial.

	Percentage of Information Transmitted		
Feature	Strategy A	Strategy B	Strategy C
Manner of articulation features			
nasal	53.14%	52.87%	59.23%
continuant	39.94%	38.42%	43.86%
voiced	58.63%	61.18%	60.38%
sibilant	62.29%	64.50%	63.28%
duration	60.14%	64.63%	63.83%
Place of articulation features			
anterior	30.64%	31.39%	28.36%
high	39.37%	39.68%	36.86%
back	36.67%	34.11%	31.62%
distributed	44.37%	45.18%	40.41%

FIGURES

Figure 1: Consonant results for each of the three rates of stimulation for each patient. The results for the different speakers and test repetitions have been combined to create this graph. Percentage correct scores are given with standard deviations. The right-most results are the average scores over all patients.

Figure 2: Vowel results for each of the three rates of stimulation for each patient. The results for the different speakers and test repetitions have been combined to create this graph. Percentage correct scores are given with standard deviations. The right-most results are the average scores over all patients.

Figure 3: Consonant results for each speaker for each patient. The results for the different strategies and test repetitions have been combined to create this graph. Percentage correct scores are given with standard deviations. The right-most results are the average scores over all patients.

Figure 4: Percentage of information transmitted for a number of phonetic distinctive features for consonant phonemes. The features used are those that showed some significant effects of strategy when tested with analysis if variance. The confusion matrices for the different patients and speaker have been combined and only data from the second trial have been included.

Figure 1: Mean consonant percent correct scores over all trials for each patient and overall mean.

	250 cycles/s	St Dev	807 cycles/s	St Dev	615 cycles/	St Dev
Patient 1	41.9270833	7.00838618	42.8385417	5.64009036	34.8958333	9.39427032
Patient 2	73.046875	5.86641167	72.65625	6.79882036	69.4010417	5.19964553
Patient 3	58.984375	5.5755932	55.2083333	7.64519941	59.375	6.71854812
Patient 4	54.8177083	6.74943199	54.5572917	7.18548684	47.9166667	9.88943507
Patient 5	59.5052083	5.3233821	57.1614583	5.48402299	59.375	4.16666667
Average	57.65625	11.6857467	56.484375	11.5448819	54.1927083	13.9084092

Figure 2: Mean vowel percent correct scores over all trials for each patient and overall mean.

	250 cycles/s	St Dev	807 cycles/s	St Dev	615 cycles/	St Dev
Patient 1	49.8355263	9.48677215	51.6447368	10.3883502	48.0263158	7.47417839
Patient 2	92.5986842	4.10152214	90.9539474	5.17190632	85.1973684	7.3731188
Patient 3	76.1513158	7.09184894	74.3421053	7.89473684	73.1907895	7.70046047
Patient 4	66.1184211	8.58798362	65.2960526	12.8640433	61.5131579	12.0737576
Patient 5	78.6184211	7.49730638	72.3684211	8.59470085	77.1381579	6.97699359
Average	72.6644737	16.0891708	70.9210526	15.7666947	69.0131579	15.480969

Figure 3: Mean consonant percent correct scores over all trials for each patient and overall mean.

	Speaker 1 (	SD	Speaker 2 (f	SD
Patient 1	42.1006944	5.89188929	37.6736111	9.57499375
Patient 2	75.2604167	5.57368908	68.1423611	4.27747436
Patient 3	57.5520833	5.69095458	58.1597222	7.91324933
Patient 4	57.3784722	7.74604384	47.4826389	6.11197739
Patient 5	59.1145833	3.77770063	58.2465278	6.09652031
Average	58.28125	12.0285223	53.9409722	12.5518179

Figure 4: Information transmission analyses

	250 cycles/s	**807 cycles/s**	**615 cycles/s**
Nasal	53.14	52.87	59.23
Continua	39.94	38.42	43.86
Voicing	58.63	61.18	60.38
Sibilant	62.29	64.5	63.28
Duration	60.14	64.63	63.83
Anterior	30.64	31.39	28.36
High	39.37	39.68	36.86
Back	36.67	34.11	31.62
Distribute	44.37	45.18	40.41

FILE MAINTENANCE AUDIT TRAIL

PURGE N

CODES -/ NAME ----------------------------------/ INITIALS --/ TITLE ---/
MATTER DESCRIPTION ------------------/ CLIENT REFERENCE -----/ OPEN LAST-MOD CLOSED WKT PACT PRES M B T 2nd&3rd NAME

MATTER HEADER INFORMATION 27OCT98 ***< WAS >***

5683 Aquatic Rentals Pty Ltd
29389 AUSTRALIA TRADE MARK APPLN   AQUATIC RENTALS AND   17SEP98 17SEP98 N .... ATA PCD PCD C 1 1 0 0
AQUATIC RENTALS PTY LTD   SP   1 1 1
AQUATIC RENTALS AND LOGO   8224

MATTER HEADER INFORMATION 27OCT98 ***< IS >***

5683 Aquatic Rentals Pty Ltd
29389 AUSTRALIA TRADE MARK APPLN NO 774669   AQUATIC RENTALS AND   17SEP98 27OCT98 N .... ATA PCD PCD C 1 1 0 0
AQUATIC RENTALS PTY LTD   SP   1 1 1
AQUATIC RENTALS AND LOGO   8224

MATTER HEADER INFORMATION 27OCT98 ***< WAS >***

5395 MT-CUP Corporation Pty Ltd
29606 UNITED STATES DESIGN APPLN NO 1052/1998   A DISPENSER   08OCT98 27OCT98 N .... ODA DAA DAA C 1 1 0 0
MT-CUP CORPORATION PTY LTD   SP   1 1 1
A DISPENSER   8224

MATTER HEADER INFORMATION 27OCT98 ***< IS >***

5395 MT-CUP Corporation Pty Ltd
29606 UNITED STATES DESIGN APPLN   A DISPENSER   08OCT98 27OCT98 N .... ODA DAA DAA C 1 1 0 0
MT-CUP CORPORATION PTY LTD   SP   1 1 1
A DISPENSER   8224

MATTER HEADER INFORMATION 27OCT98 ***< WAS >***

5395 MT-CUP Corporation Pty Ltd
29605 UNITED STATES DESIGN APPLN NO 1053/1998   A COLLAR   08OCT98 27OCT98 N .... ODA DAA DAA C 1 1 0 0
MT-CUP CORPORATION PTY LTD   SP   1 1 1
A COLLAR   8224

MATTER HEADER INFORMATION 27OCT98 ***< IS >***

5395 MT-CUP Corporation Pty Ltd
29605 UNITED STATES DESIGN APPLN   A COLLAR   08OCT98 27OCT98 N .... ODA DAA DAA C 1 1 0 0
MT-CUP CORPORATION PTY LTD   SP   1 1 1
A COLLAR   8224